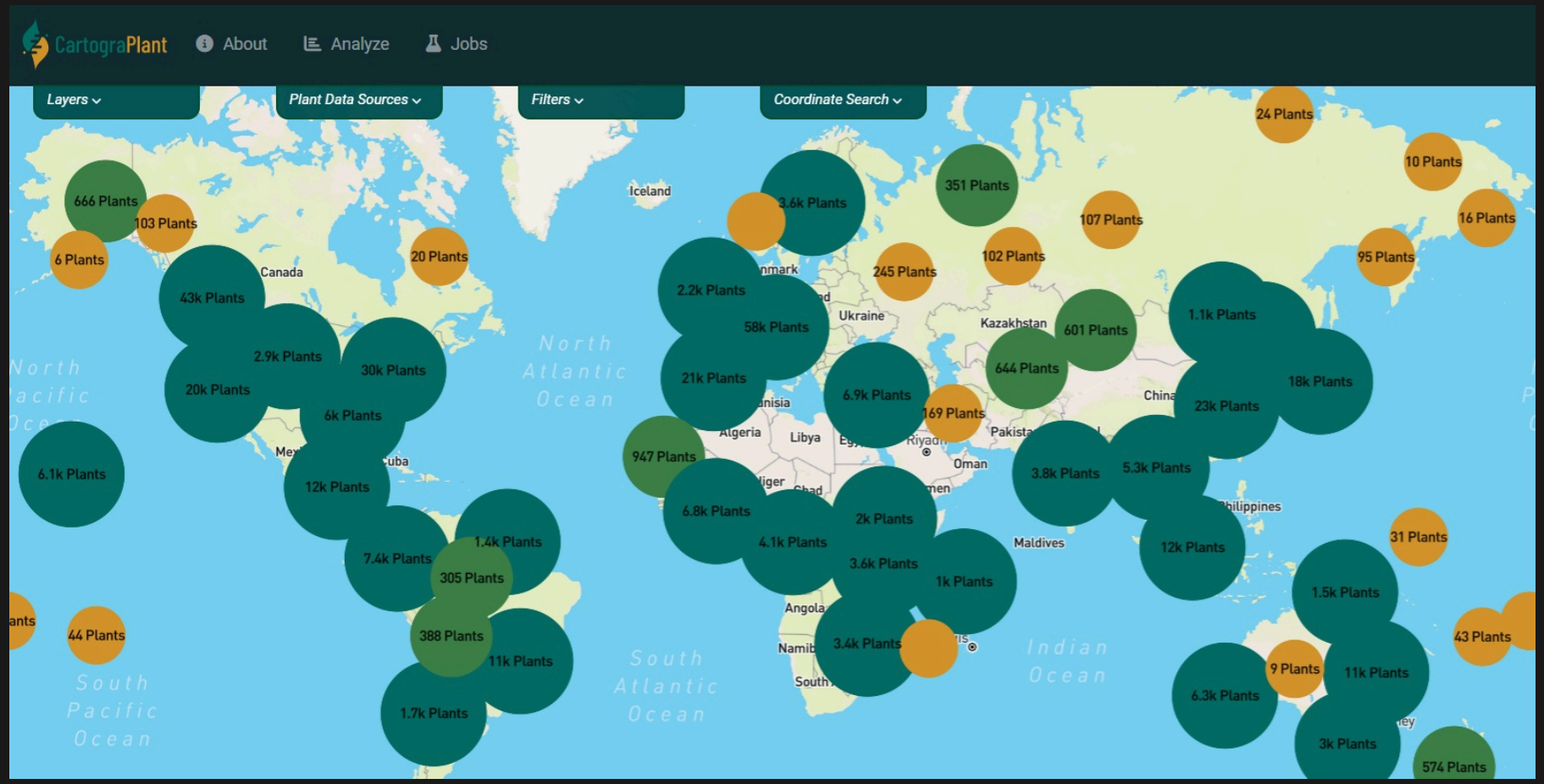
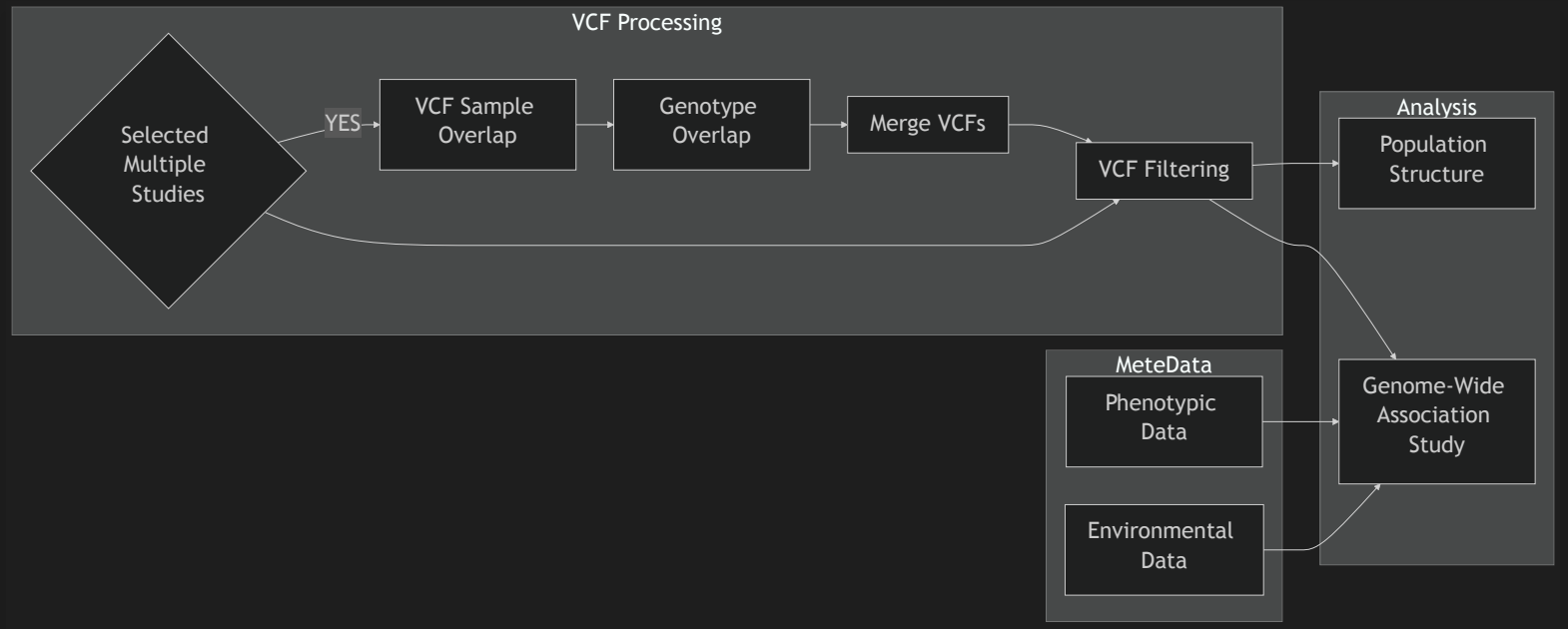
| Analysis Panel | Landscape Genomics | Demo |



ANALYSIS PANEL OVERVIEW



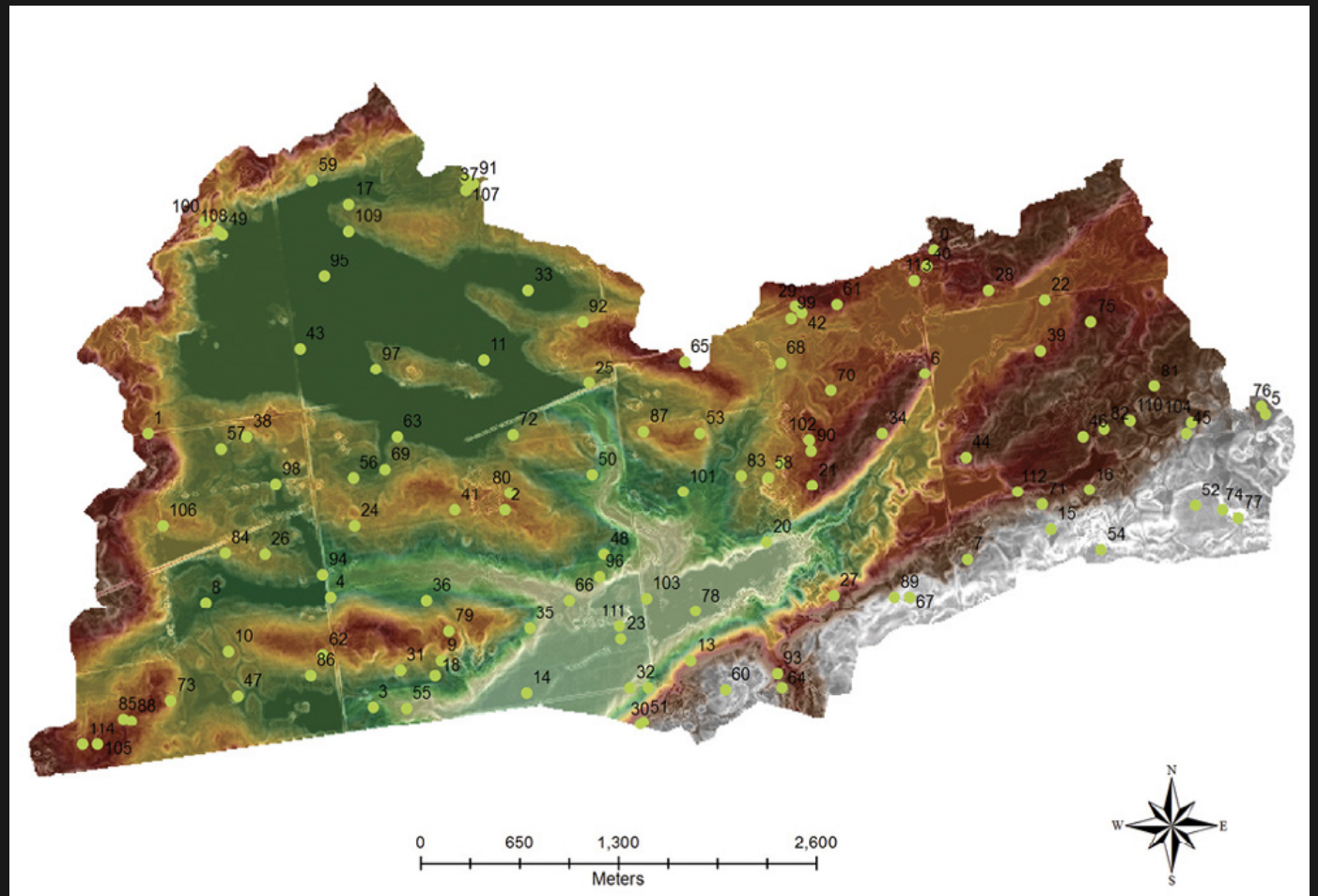
NEXTFLOW

Workflow management software

- Reproducible
- Concurrency
- Portable
- nf-core Community

LANDSCAPE GENOMICS

Aims to understand how environmental heterogeneity influences the distribution of genetic diversity.



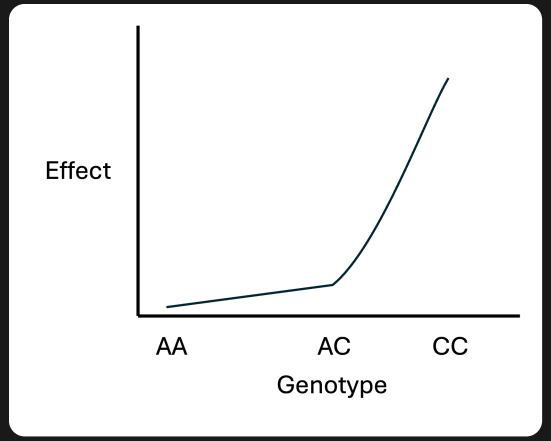
POPULATION STRUCTURE

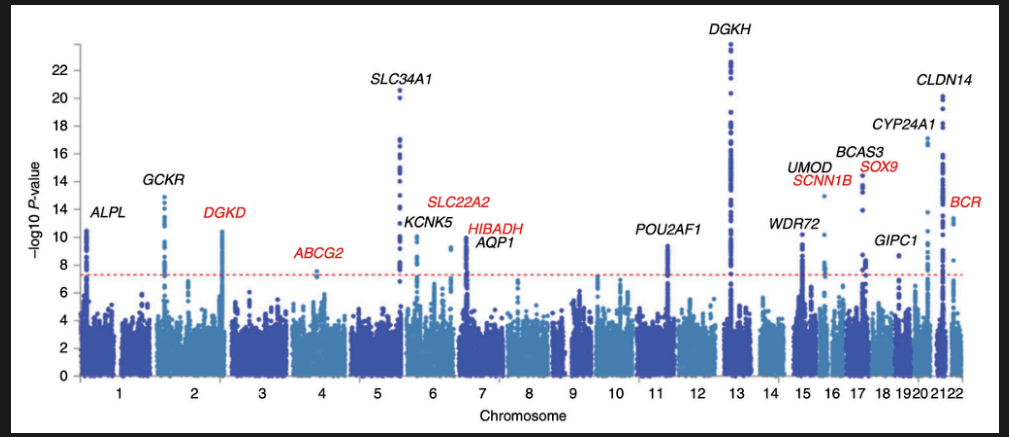
Patterns of genetic variation within and between individuals

- Isolation by distance
- Genetic drift
- Selection

GENOME-WIDE ASSOCIATION STUDY (GWAS)

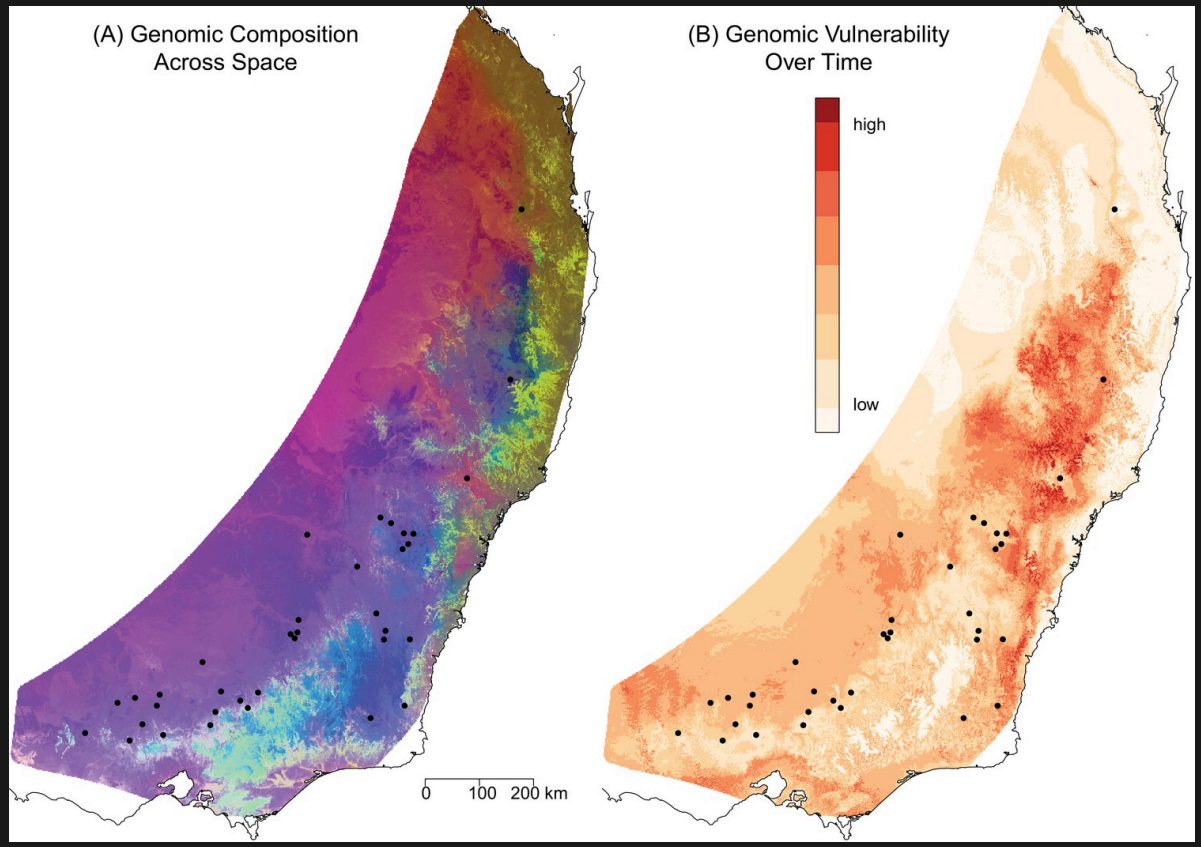
Examines the relationship between different genotypes and specific traits or environments.





GENETIC OFFSET

Difference between the current genetic makeup to the genetic makeup predicted to be optimal for future environmental conditions



TPPS SUBMISSION ENABLES SEAMLESS INTEGRATION OF STUDIES

Allowing for analysis of large datasets in the analysis panel

Mega:
Integration of raw data

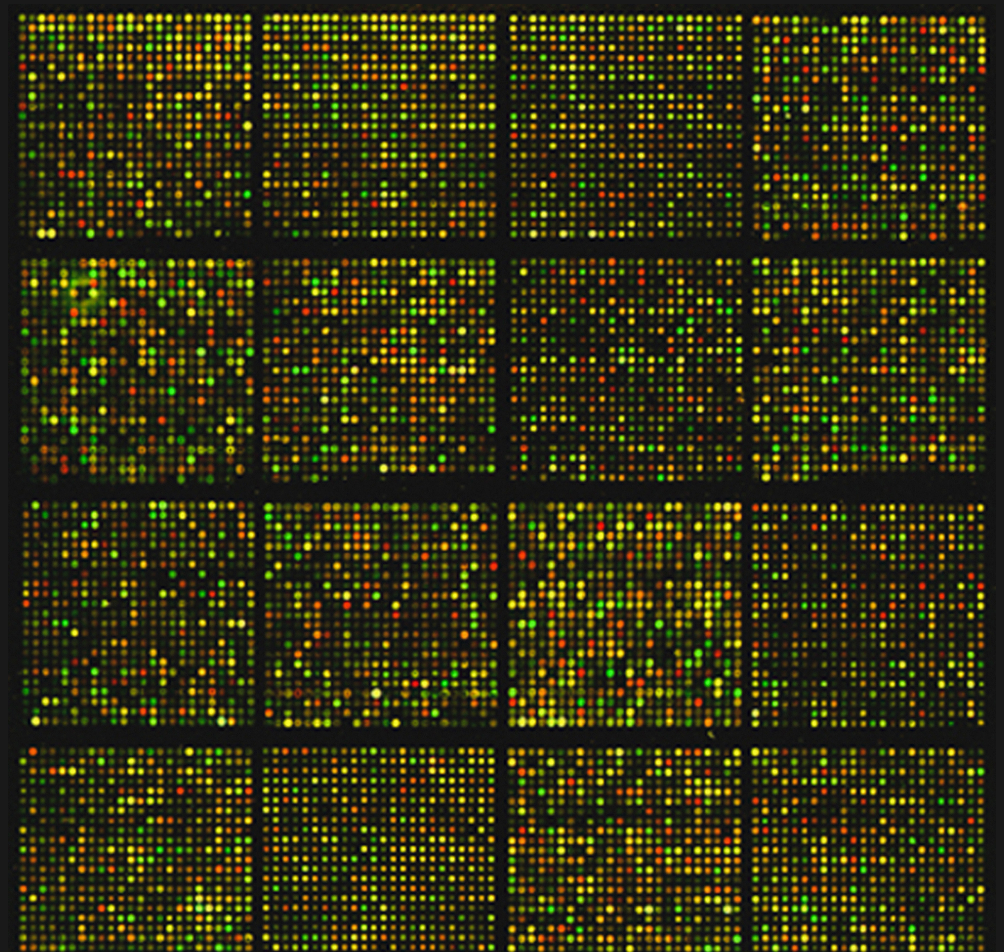
Meta:
Integration of summary results

DEMO WITH POPULUS TRICHOCARPA

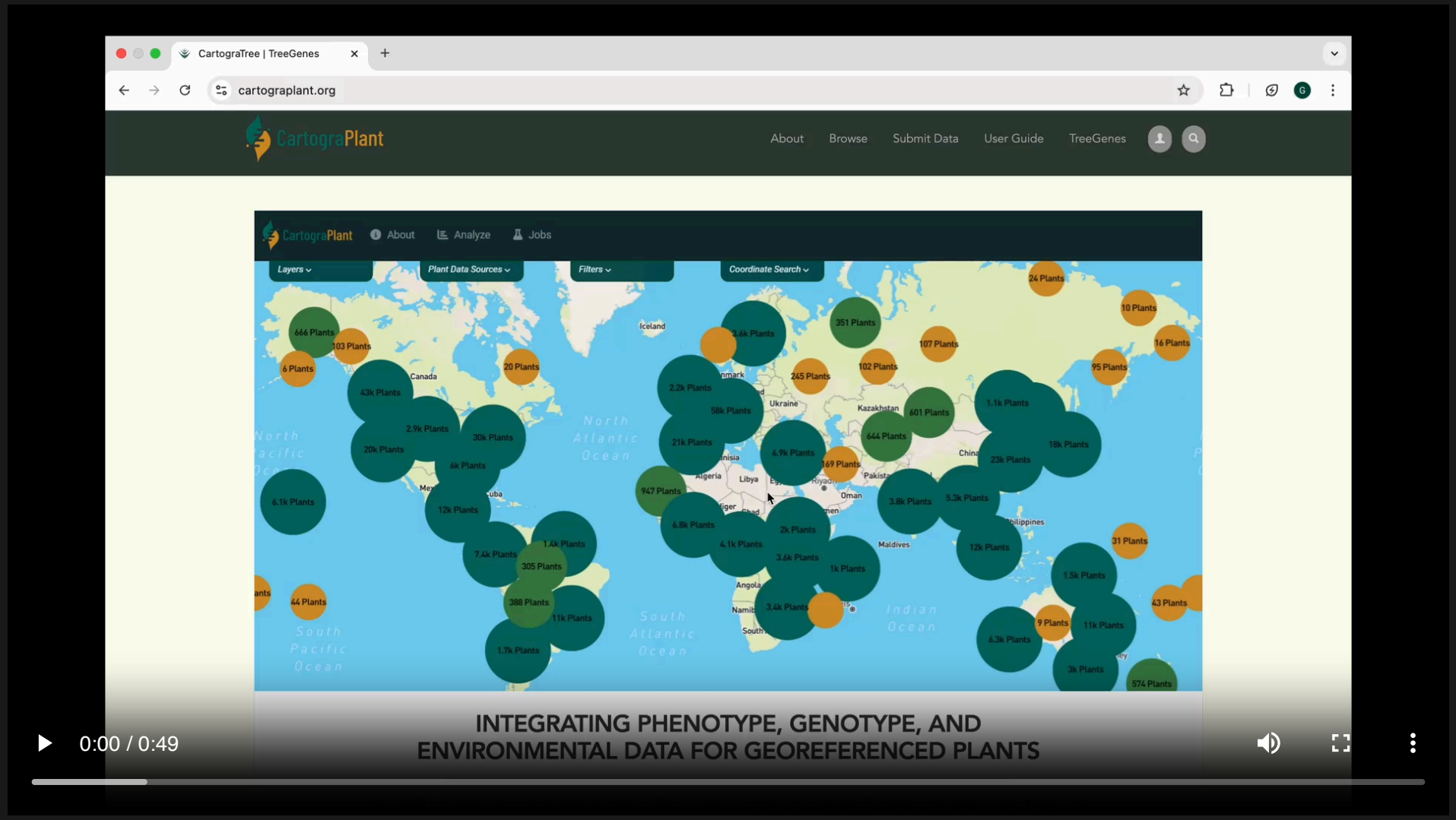
Study	Samples	Genotypes	Phenotypes
Geraldes et al. 2008	55	32,000	0
Mckown et al. 2013	555	0	400
Mckown et al. 2014	555	28,000	0

GENOTYPE ASSAY

A cost effective way to obtain tens of thousands of mutations



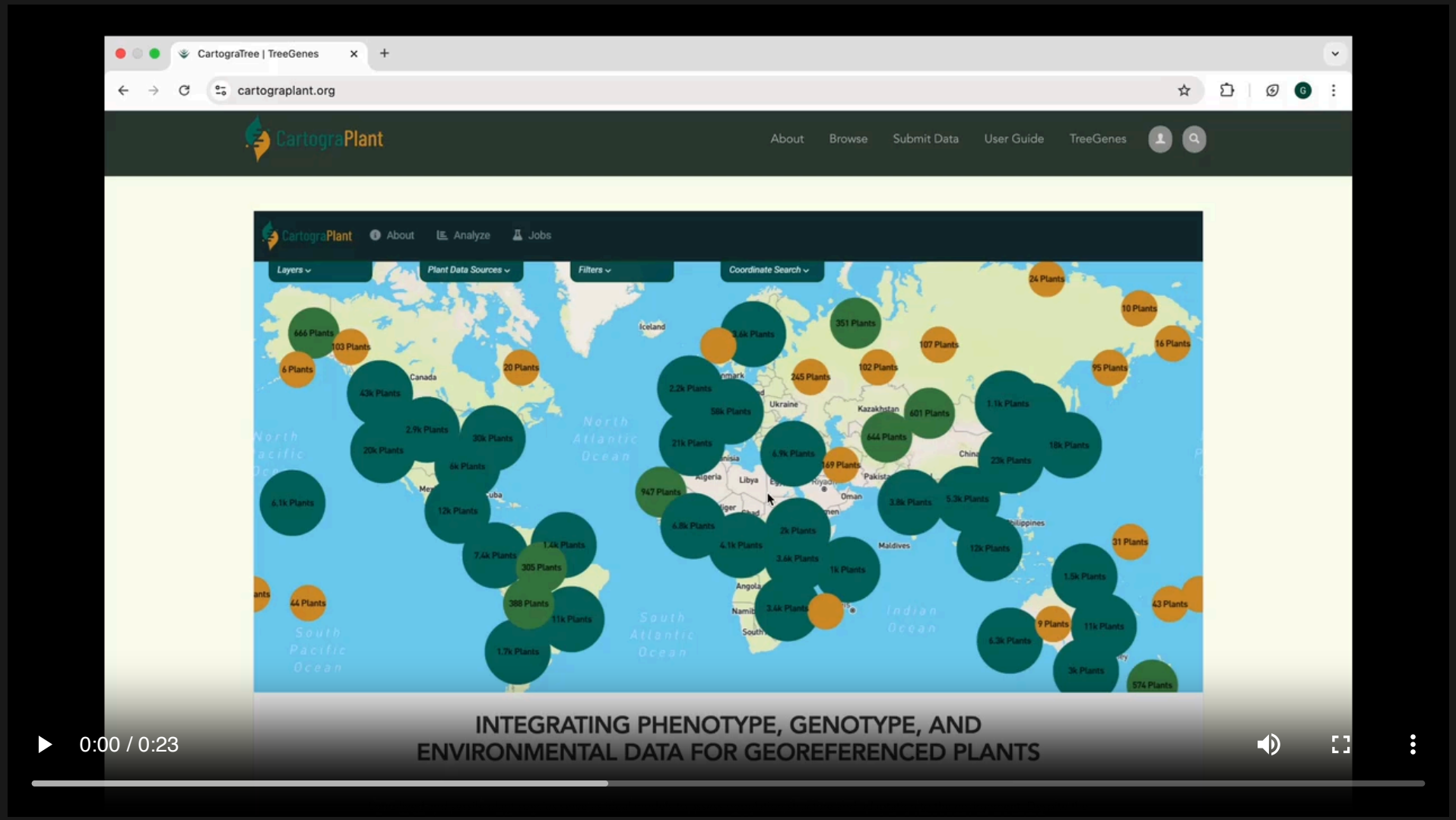
FILTER AND SELECT STUDIES IN CARTOGRAPLANT



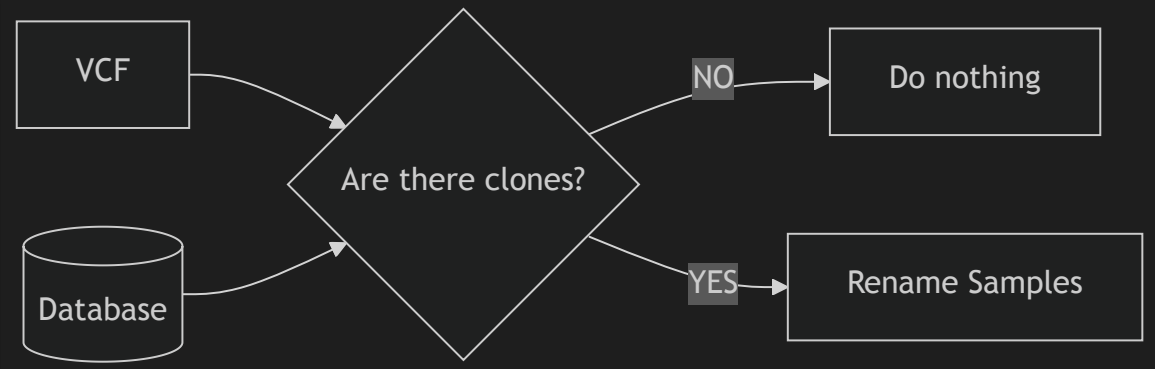
PANEL 1: WORKSPACE

Responsible for file management:

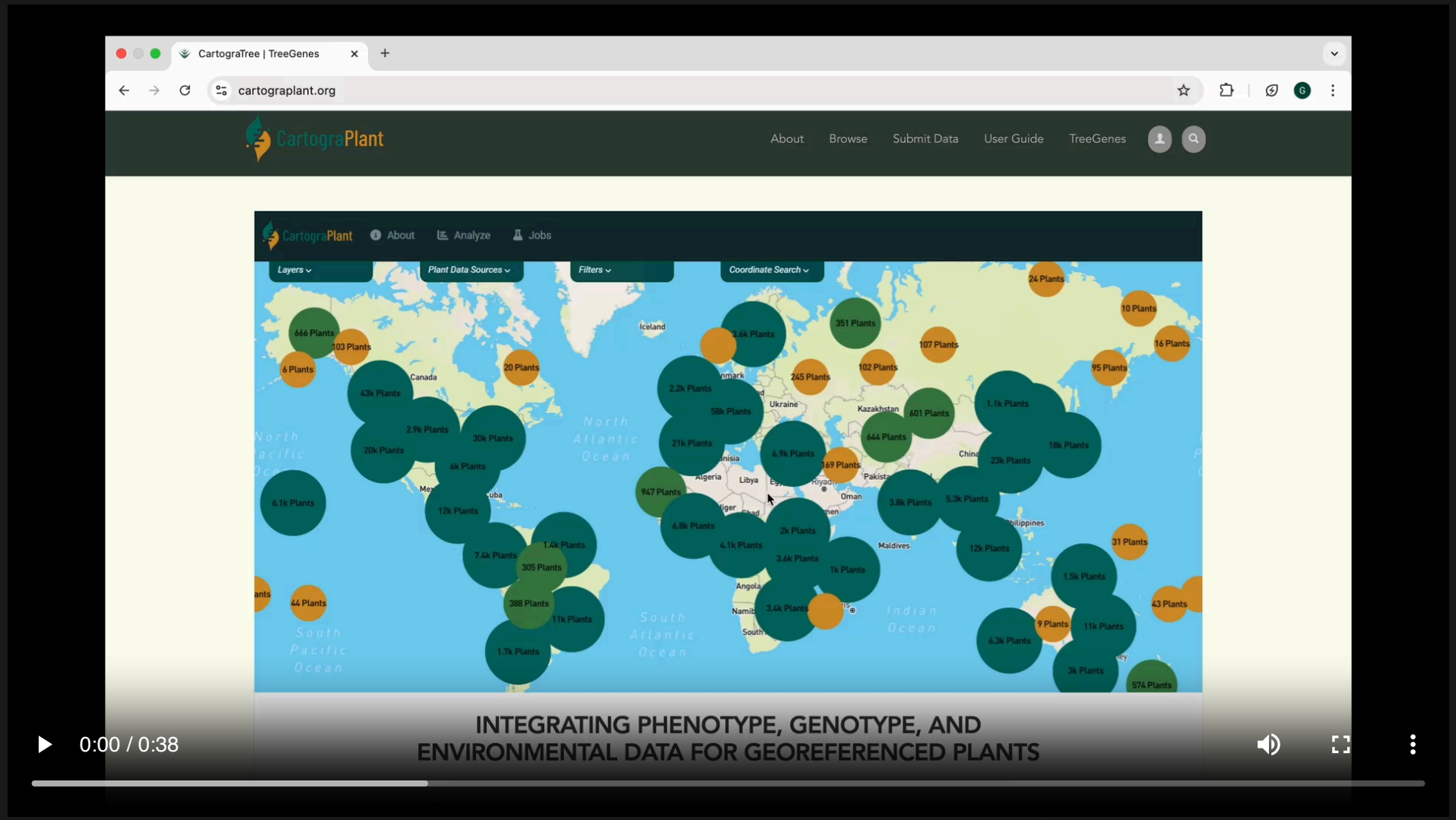
- Create new workspace directory
- Upload files from local machine
- Stores output from workflows
- Delete Files


CartograTree | TreeGenes
cartograplant.org
CartograPlant
About
Browse
Submit Data
User Guide
TreeGenes
CartograPlant
About
Analyze
Jobs
Layers
Plant Data Sources
Filters
Coordinate Search
INTEGRATING PHENOTYPE, GENOTYPE, AND
ENVIRONMENTAL DATA FOR GEOREFERENCED PLANTS
0:00 / 0:23

PANEL 2: STUDY-CONTEXT



Study 1	Study 2	Study 3	Standard Name
GLCA-26-1	GLCA26-1	GLCA26-1	sample05
HFCA-20-2	Null	Null	sample06


CartograTree | TreeGenes
cartograplant.org
CartograPlant
About
Browse
Submit Data
User Guide
TreeGenes
CartograPlant
About
Analyze
Jobs
Layers
Plant Data Sources
Filters
Coordinate Search
INTEGRATING PHENOTYPE, GENOTYPE, AND
ENVIRONMENTAL DATA FOR GEOREFERENCED PLANTS
0:00 / 0:38

PANEL 3: FILTER BY TRAITS

Select phenotypes of interest for upload into workspace

- Determine total counts
- Redundancy analysis based on x>1 phenotypes

CartograTree | TreeGenes

treegenesdb.org/cartogratree#

CartograPlant

Analysis ID	Studies	Plants	Species	Phenotypes	Genotypes	Environmental layers
3541	3	963	1	2	28083	2

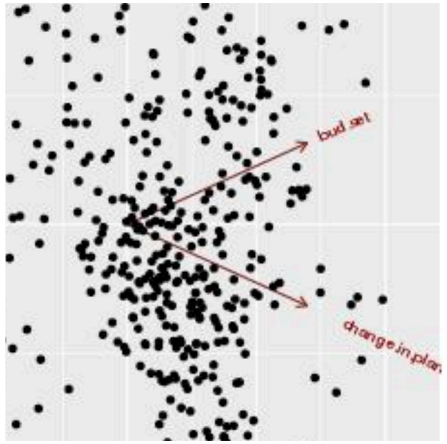
- Manage analysis
- Study context
- **Filter By Traits**
- Study markers overlap
- Filtering & Imputation
- Population Structure
- Environmental Data
- Run Analysis
- Summary and Confirm

3 studies based on analysis study context choices

TGDR1892 TGDR1904 TGDR2269

☐ bole density	519 phenotypes	⛔	No overlaps
☐ bole mass	519 phenotypes	⛔	No overlaps
☐ bud break	916 phenotypes	⛔	No overlaps
☑ bud set	1,370 phenotypes	⛔	No overlaps
☐ canopy duration	910 phenotypes	⛔	No overlaps
☑ change in plant height	1,368 phenotypes	⛔	No overlaps
☐ change in plant volume	916 phenotypes	⛔	No overlaps
☐ day 100% of leaves are yellow	455 phenotypes	⛔	No overlaps
☐ day 25% of leaves are yellow	458 phenotypes	⛔	No overlaps
☐ day 50% of leaves are yellow	458 phenotypes	⛔	No overlaps
☐ day 75% of leaves are yellow	458 phenotypes	⛔	No overlaps
☐ elevation	443 phenotypes	⛔	No overlaps
☐ frost free days	443 phenotypes	⛔	No overlaps
☐ leaf chlorophyll content, spring	203 phenotypes	⛔	No overlaps
☐ leaf chlorophyll content, summer	777 phenotypes	⛔	No overlaps
☐ leaf flush	1,571 phenotypes	⛔	No overlaps
☐ leaf lifespan	458 phenotypes	⛔	No overlaps
☐ leaf shape	458 phenotypes	⛔	No overlaps
☐ log(plant height growth)	457 phenotypes	⛔	No overlaps
☐ log(plant volume growth)	457 phenotypes	⛔	No overlaps
☐ number of branches	458 phenotypes	⛔	No overlaps
☐ plant height	1,826 phenotypes	⛔	No overlaps
☐ plant height growth cessation date	458 phenotypes	⛔	No overlaps
☐ whole plant mass	519 phenotypes	⛔	No overlaps

PCA (2 phenotypes)



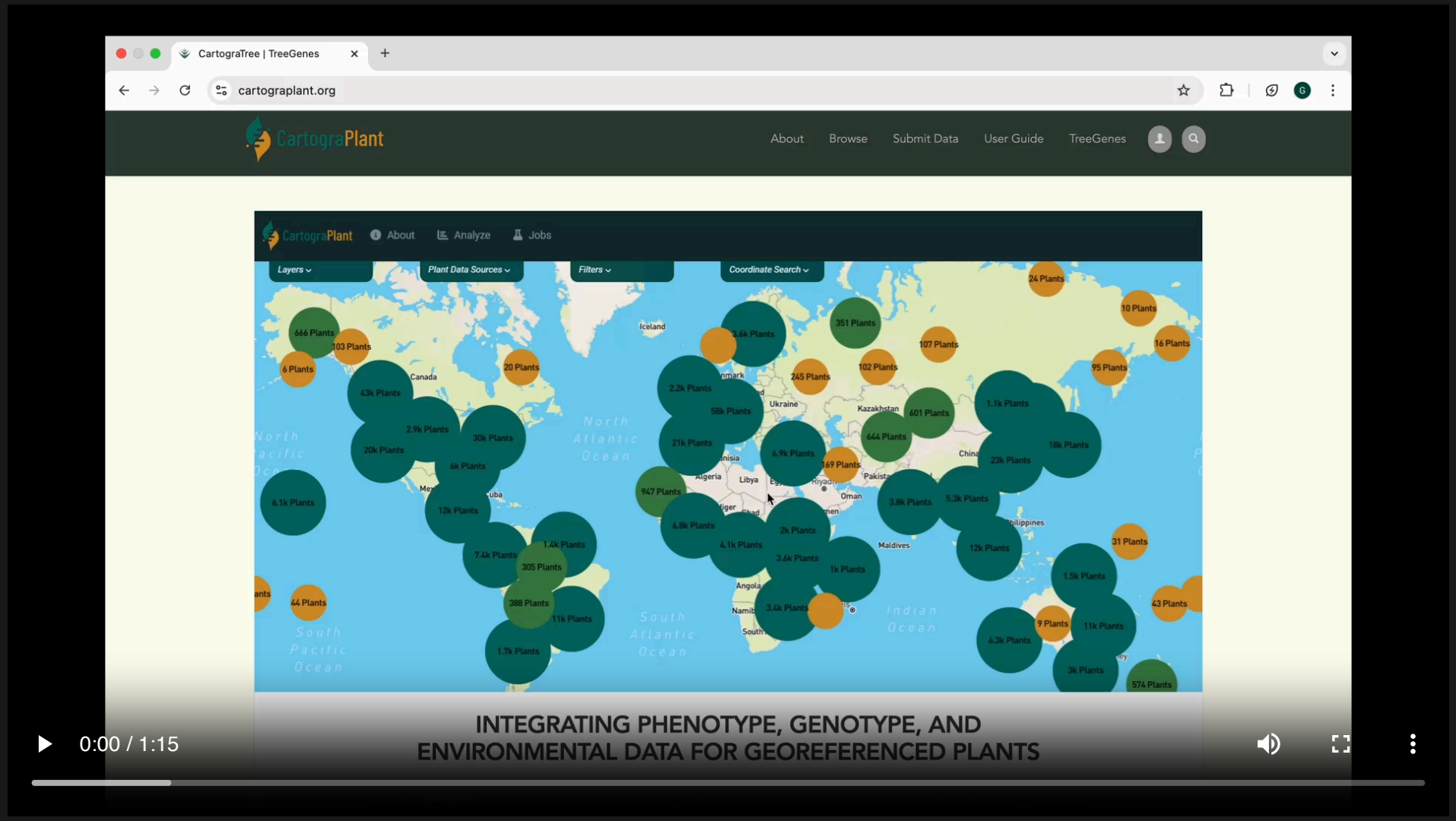
PHENOTYPE FILE OUTPUT FORMAT

phenotype_id	phenotype_name	plant_accession	study_accession	year	value
8880144	change in plant height	TGDR1904-ALAA20-1	TGDR1904	2009	194.7

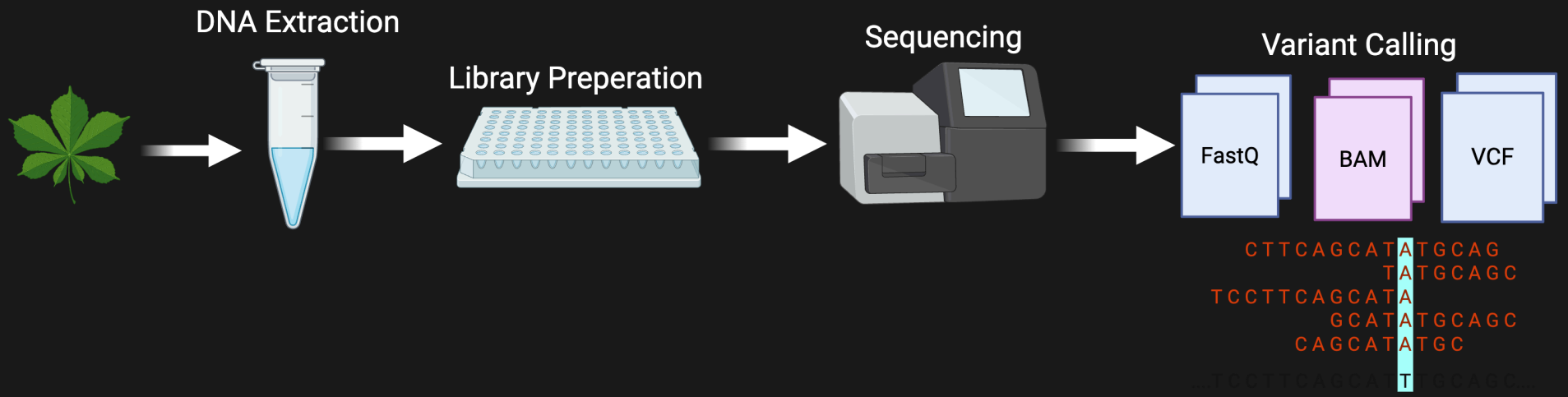
PANEL 4: STUDY MARKERS OVERLAP

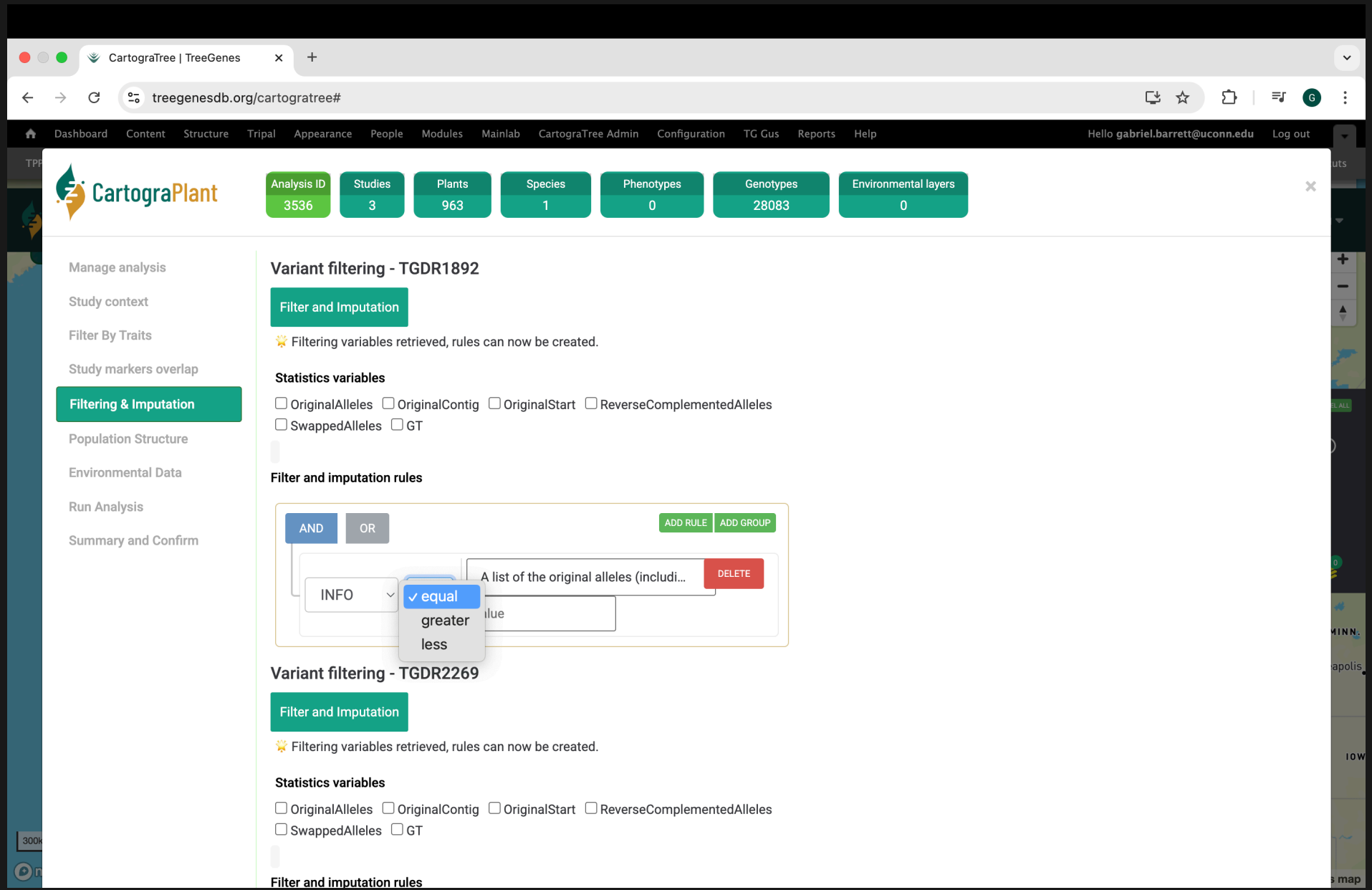
1. Marker Overlap
2. Genotype concordance
3. Merge VCF's

GENOTYPE QUALITY CONTROL

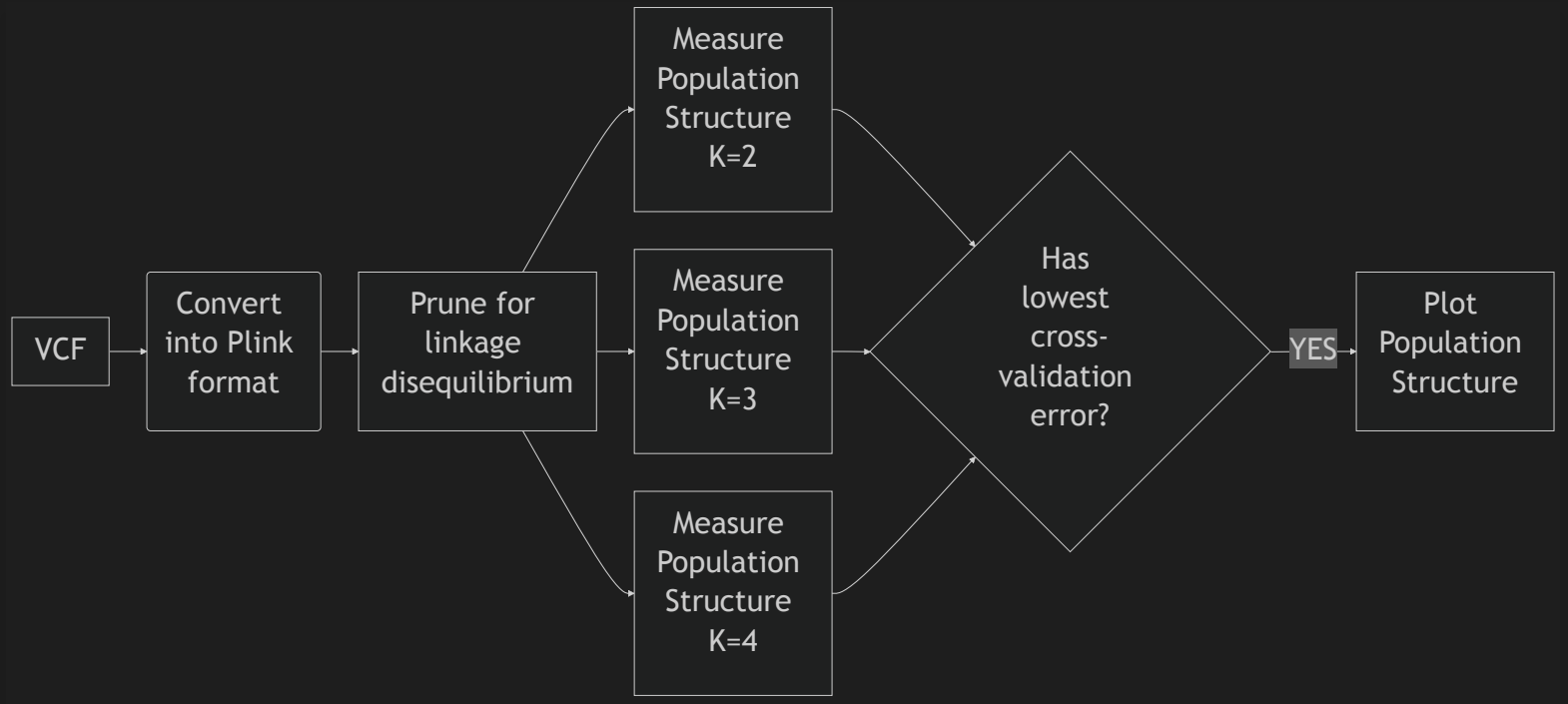


PANEL 5: VARIANT FILTERING



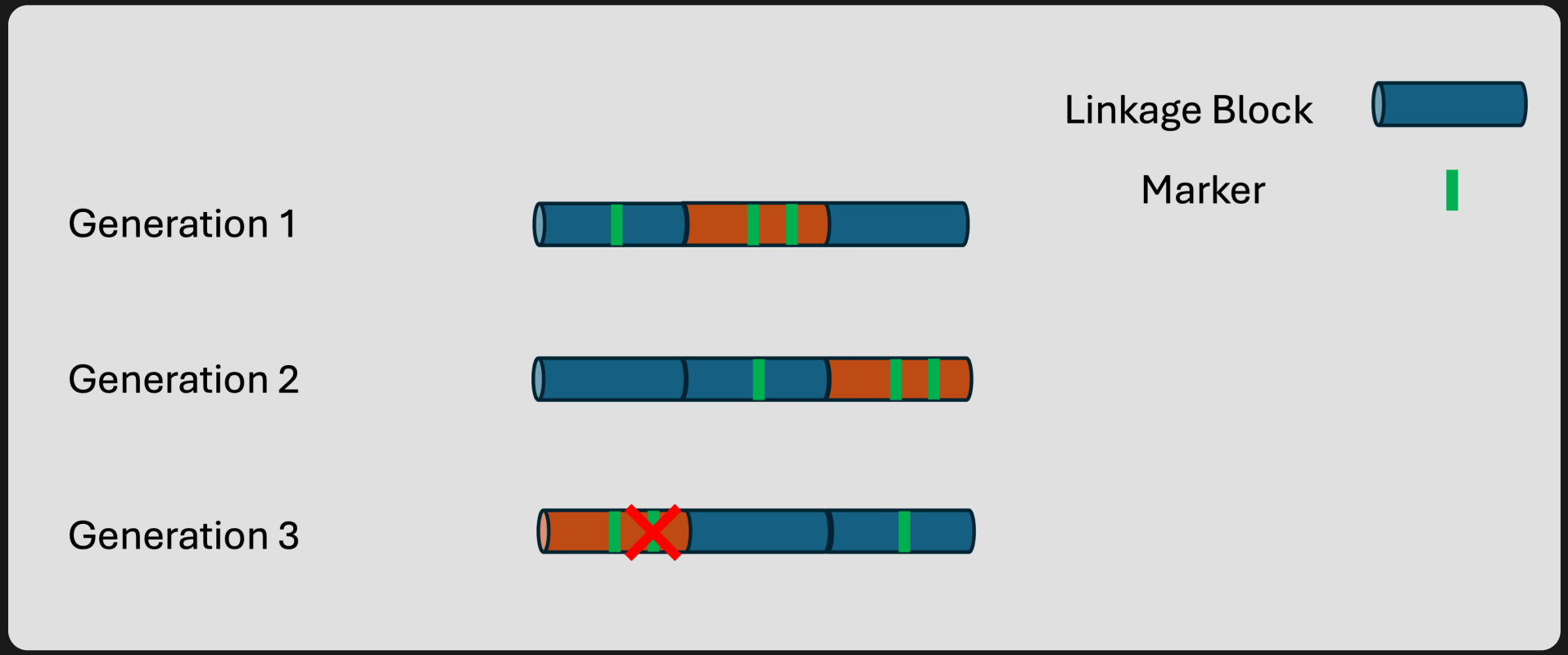

CartograTree | TreeGenes
treegenesdb.org/cartogratree#
Dashboard
Content
Structure
Tripal
Appearance
People
Modules
Mainlab
CartograTree Admin
Configuration
TG Gus
Reports
Help
Hello gabriel.barrett@uconn.edu
Log out
CartograPlant
Analysis ID
3536
Studies
3
Plants
963
Species
1
Phenotypes
0
Genotypes
28083
Environmental layers
0
Manage analysis
Study context
Filter By Traits
Study markers overlap
Filtering & Imputation
Population Structure
Environmental Data
Run Analysis
Summary and Confirm
Variant filtering - TGDR1892
Filter and Imputation
Filtering variables retrieved, rules can now be created.
Statistics variables
OriginalAlleles
OriginalContig
OriginalStart
ReverseComplementedAlleles
SwappedAlleles
GT
Filter and imputation rules
AND
OR
ADD RULE
ADD GROUP
INFO
A list of the original alleles (includi...
DELETE
equal
greater
less
Variant filtering - TGDR2269
Filter and Imputation
Filtering variables retrieved, rules can now be created.
Statistics variables
OriginalAlleles
OriginalContig
OriginalStart
ReverseComplementedAlleles
SwappedAlleles
GT
Filter and imputation rules

PANEL 6: POPULATION STRUCTURE



LINKAGE DISEQUILIBRIUM

Non-random associate of alleles at different loci on a chromosome



CartograTree | TreeGenes

treegenesdb.org/cartogratree#

Dashboard Content Structure Tripal Appearance People Modules Mainlab CartograTree Admin Configuration TG Gus Reports Help Hello gabriel.barrett@uconn.edu Log out

CartograPlant

Analysis ID	Studies	Plants	Species	Phenotypes	Genotypes	Environmental layers
3526	3	963	2	0	28083	2

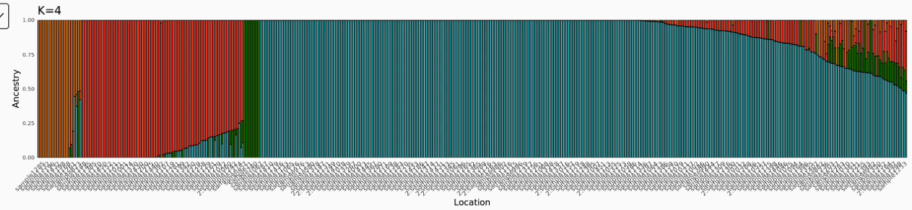
- Manage analysis
- Study context
- Study markers overlap
- Filtering & Imputation
- **Population Structure**
- Environmental Data
- Run Analysis
- Summary and Confirm

Nextflow Population Structure Workflow

Path To Vcf File

Geraldes_2013.v2_TGDR1892_Potr.2_0-Potr.4_1.vcf.vcf.gz

Select appropriate variant call file



Model Complexity

3,4,5

Window Size To Calculate Linkage Disequilibrium

150

Number Of Variants To Consider In A Window

10

The Correlation Between A Pair Of Loci

0.2

How To Handle Partially Missing Genotypes When Converting File

Treat half-calls as haploid/homozygous

How To Handle Converting VCF ID's To Plink IID And FIID

0

GENERATE POPULATION STRUCTURE

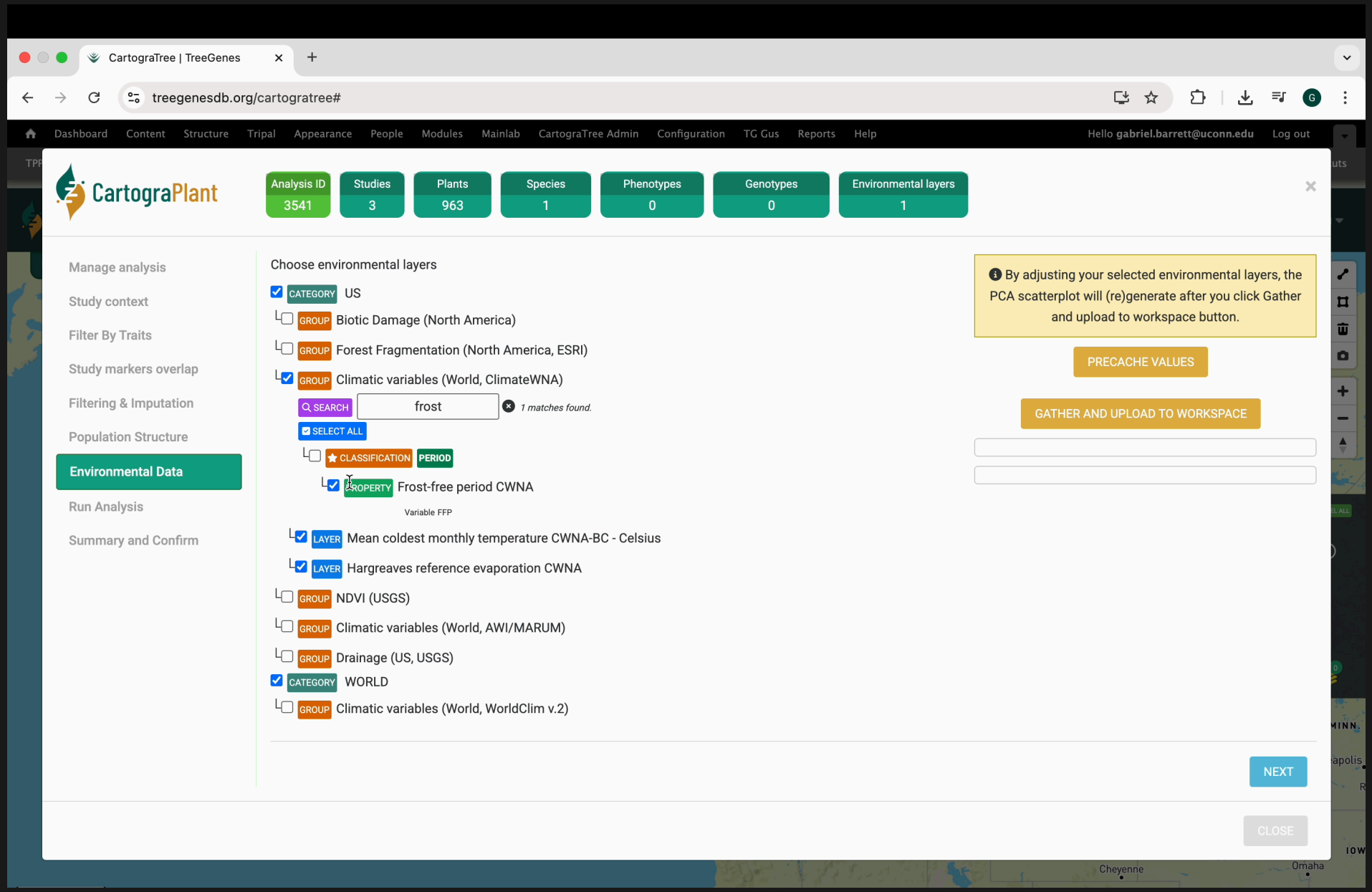
NEXT

CLOSE

https://treegenesdb.org/cartogratree#analysis-popstruct-section

PANEL 7: CHOOSE ENVIRONMENTS

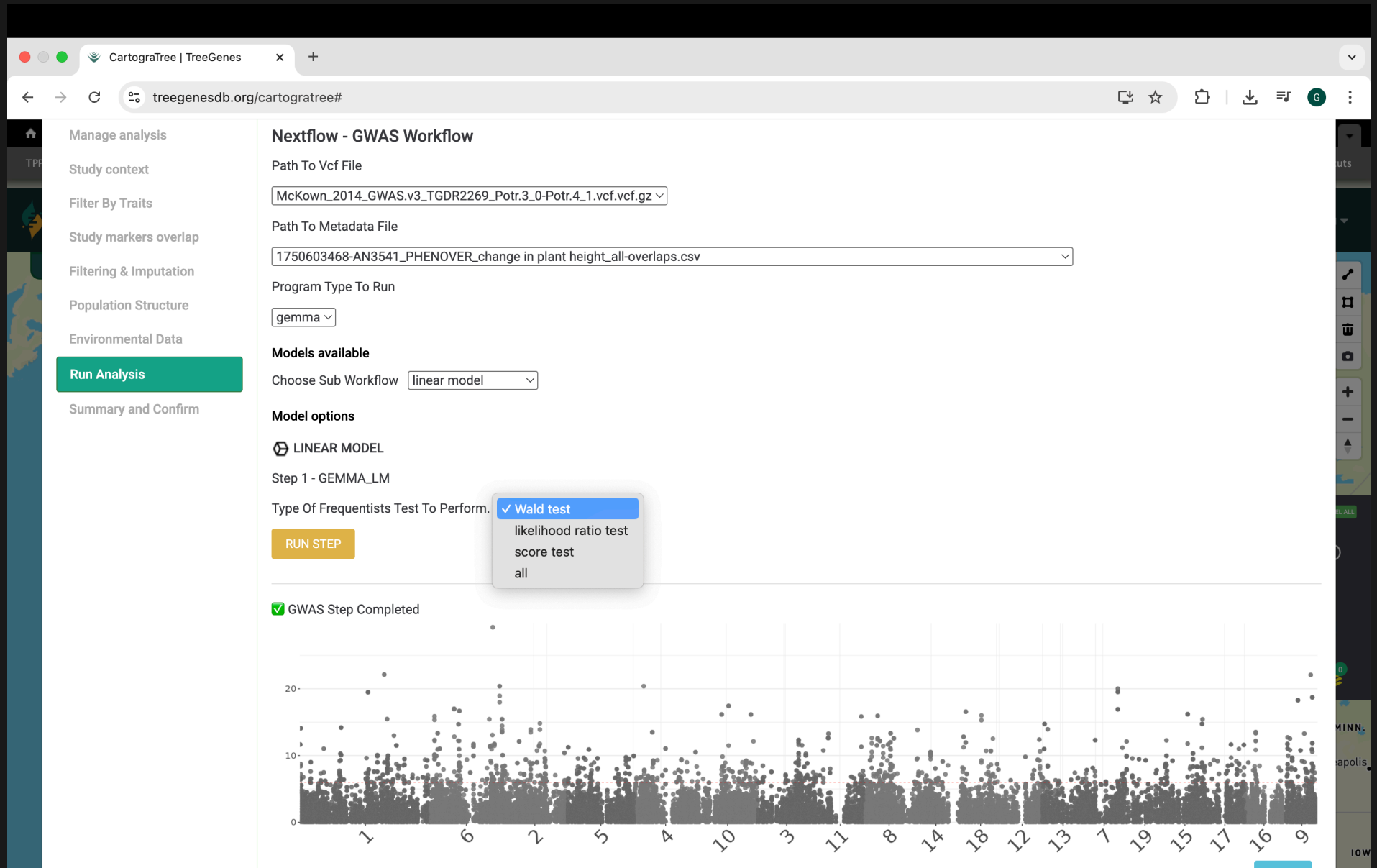
multicollinearity: describes a high correlation of two or more independent variables


CartograTree | TreeGenes
treegenesdb.org/cartogratree#
Dashboard Content Structure Tripal Appearance People Modules Mainlab CartograTree Admin Configuration TG Gus Reports Help
Hello gabriel.barrett@uconn.edu Log out
CartograPlant
Analysis ID 3541
Studies 3
Plants 963
Species 1
Phenotypes 0
Genotypes 0
Environmental layers 1
Manage analysis
Study context
Filter By Traits
Study markers overlap
Filtering & Imputation
Population Structure
Environmental Data
Run Analysis
Summary and Confirm
Choose environmental layers
CATEGORY US
GROUP Biotic Damage (North America)
GROUP Forest Fragmentation (North America, ESRI)
GROUP Climatic variables (World, ClimateWNA)
SEARCH frost
1 matches found.
SELECT ALL
CLASSIFICATION PERIOD
PROPERTY Frost-free period CWNA
Variable FFP
LAYER Mean coldest monthly temperature CWNA-BC - Celsius
LAYER Hargreaves reference evaporation CWNA
GROUP NDVI (USGS)
GROUP Climatic variables (World, AWI/MARUM)
GROUP Drainage (US, USGS)
CATEGORY WORLD
GROUP Climatic variables (World, WorldClim v.2)
By adjusting your selected environmental layers, the PCA scatterplot will (re)generate after you click Gather and upload to workspace button.
PRECACHE VALUES
GATHER AND UPLOAD TO WORKSPACE
NEXT
CLOSE
Cheyenne
Omaha

PANEL 8: RUN ANALYSIS

Perform G x P and G x E associations

LINEAR MODEL (LM) IN GEMMA



SPURIOUS CORRELATION

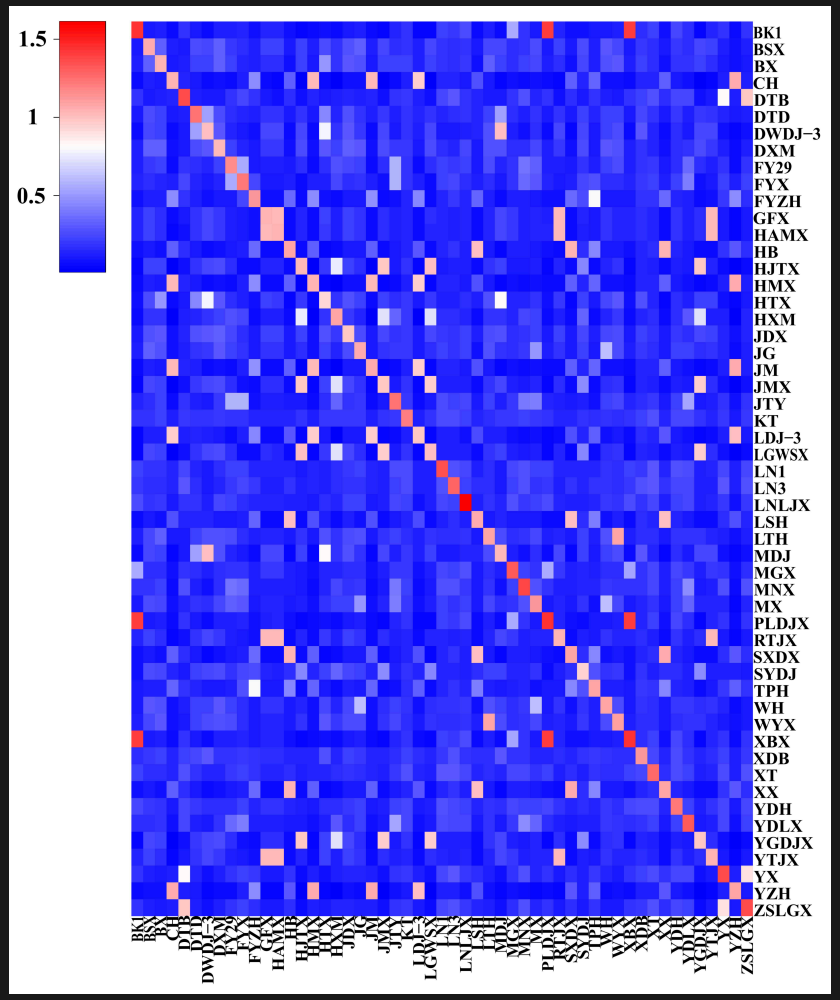
LINEAR MIXED MODEL (LMM) IN GEMMA

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$$

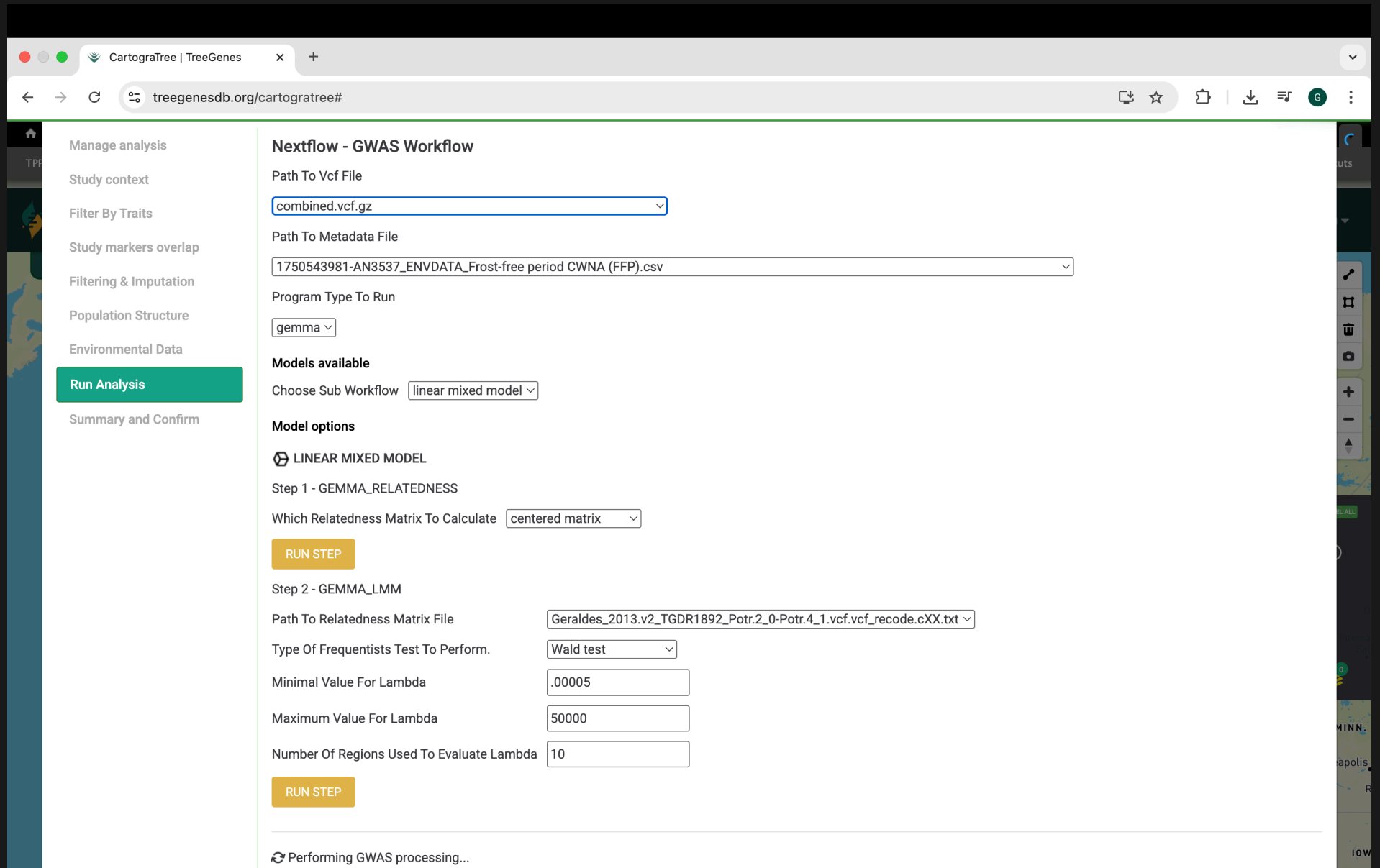
- $\mathbf{y}$: vector of phenotypes
- $\mathbf{X}$: matrix of fixed effects (e.g., SNPs, covariates)
- $\boldsymbol{\beta}$: vector of fixed-effect coefficients
- $\mathbf{Z}$: incidence matrix for random effects
- $\mathbf{u} \sim \mathcal{N}(0, \mathbf{K}\sigma_g^2)$: random genetic effects with kinship matrix $\mathbf{K}$
- $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$: residual errors

KINSHIP MATRIX

Measure the degree of relatedness between samples



LINEAR MIXED MODEL (LMM) IN GEMMA

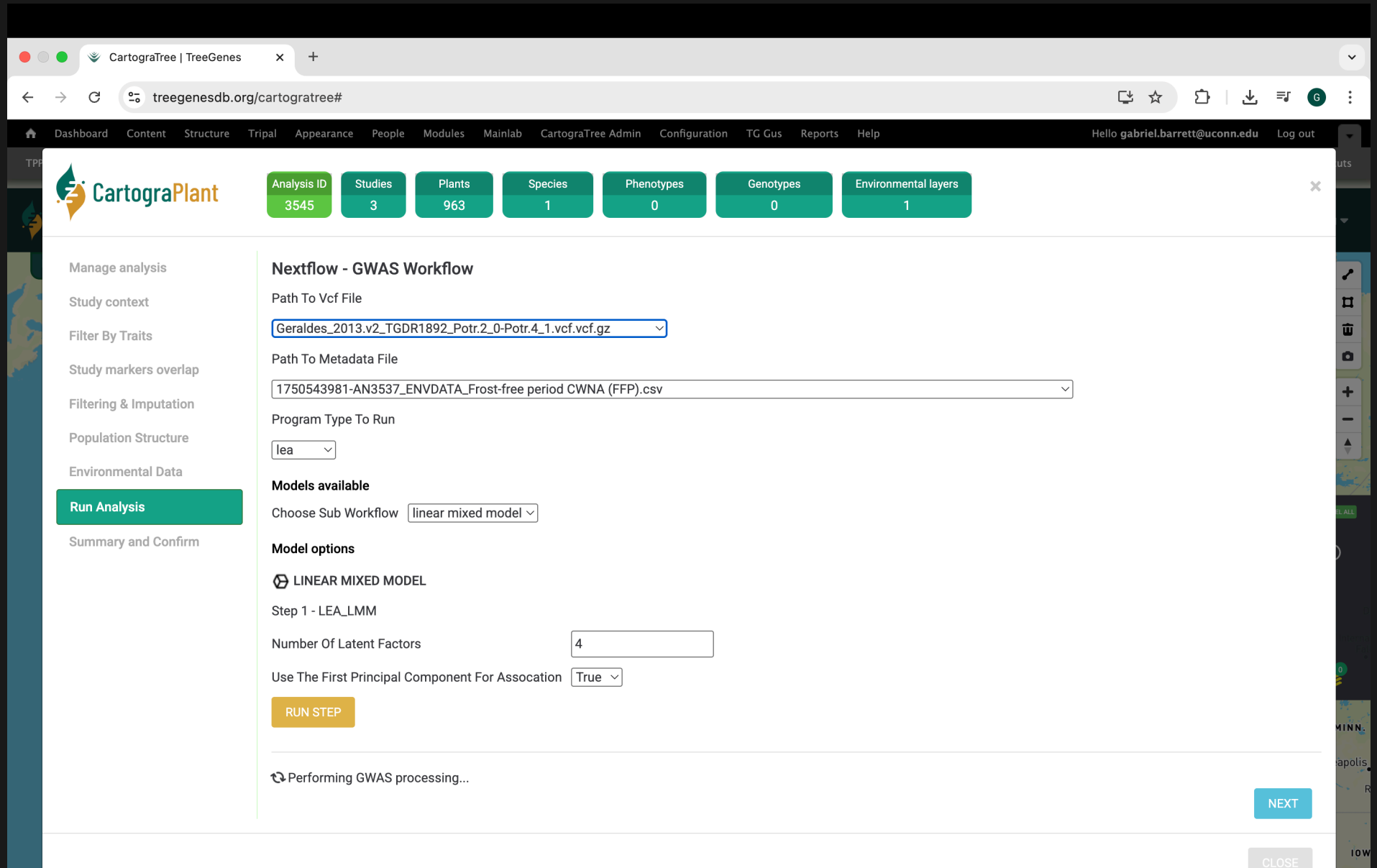


LINEAR MIXED MODEL (LMM) IN LEA

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{V}^\top + \boldsymbol{\epsilon}$$

- $\mathbf{Y}$: matrix of genotypes (individuals × loci)
- $\mathbf{X}$: matrix of environmental variables (individuals × predictors)
- $\boldsymbol{\beta}$: matrix of effect sizes (predictors × loci)
- $\mathbf{U}$: matrix of latent factors (individuals × K)
- $\mathbf{V}$: matrix of factor loadings (loci × K)
- $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)$: residual noise

LINEAR MIXED MODEL (LMM) IN LEA



Acknowledgements

Biocuration team
Isabella Harding
Trang Nguyen
Phoebe Zhou

Development team
Emily Grau
Gabe Barrett
Risharde Ramanth
Vlad Savitsky

Dr Jill Wegrzyn
Dr Margaret Staton
Dr Stephen Flickin
Dr Nic Herndon

TreeGenes
A Forest Tree Genome Database
treegenesdb.org

CartograPlant
cartograplant.org

treesnap.org

UCONN
UNIVERSITY OF CONNECTICUT

Tripal

THE UNIVERSITY OF TENNESSEE KNOXVILLE

nextflow

ECU

The Nature Conservancy

PLANT COMPUTATIONAL GENOMICS

USDA